

EGU2020-18670

<https://doi.org/10.5194/egusphere-egu2020-18670>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



Data Tailor: Integrate EUMETSAT's data into your datacube

Daniel Lee, Rodrigo Romero, Peter Miu, Fernando Jose Pereda Garcimartin, and Oscar Perez Navarro

EUMETSAT, Darmstadt, Germany (daniel.lee@eumetsat.int)

EUMETSAT hosts a large collection of geophysical data sets that have been produced by over 35 years of operational meteorological satellites. This trove of remote sensing data products is complex, featuring observations from multiple generations of polar and geostationary satellites. Each mission has different primary objectives, resulting in different instrument payloads, resolutions, and variables observed. As EUMETSAT's next-generation core missions are launched and joined by smaller missions with narrower foci, both the size and complexity of these data will increase exponentially.

The data alone are a valuable resource for the geosciences, but the value that can be extracted from them increases greatly when they are combined with data from other disciplines. As EUMETSAT's primary missions are focused on observational meteorology, the potential synergies with e.g. numerical weather prediction data are readily apparent. However, EUMETSAT data is increasingly used in applications from other domains, e.g. oceanography, agriculture, and atmospheric composition, to name just a few.

New solutions are being implemented to unlock the potential of EUMETSAT's data, particularly in combination with data from other disciplines and leveraging emerging data-driven approaches such as data mining and machine learning. A particular challenge in this regard is the heterogeneity of the individual data products, each of which is optimised to accurately describe the observed variable and quality information associated with the observing instrument and platform. A further challenge is the heterogeneity of the potential users, all of whom have preferred toolsets and processing chains.

The EUMETSAT Data Tailor is part of a larger initiative at EUMETSAT to support users in taking full advantage of our data holdings. It addresses the problem that there is no single "best format" for all users by allowing users to tailor data products to fit their needs. With it, users can extract the data that is relevant for them by selecting by geospatial and spectral criteria, resample into the projection and resolution that they require, and reformat the data into a variety of popular formats. Tailoring workflows can be created graphically or written by hand in YAML and saved in a given Data Tailor deployment.

The Data Tailor is cloud-native, exposing its functionality as a microservice, via a web UI, on the command line, and as a Python package. Support for additional functions can be added easily via

its plug-in architecture, which allows dynamically adding and removing functionality to an installation. It is released under an Apache v2 license, making it easy to deploy the software in any context. Whether data is in flight or at rest, the Data Tailor offers users easy access to EUMETSAT products in the format of their choice.

This presentation will showcase the Data Tailor and briefly address other exciting developments at EUMETSAT that the Data Tailor is integrated with that will support big data workflows with EUMETSAT's past, present, and future data.